

Pré-print du 01/10/2007 de l'article :

Kosmopoulos C., Pumain D., (2008), «Révolution numérique et évaluation bibliométrique dans les sciences humaines et sociales

La version finalisée de cet article a été publiée dans la *Revue européenne des sciences sociales (Cahiers Vilfredo Pareto)*, vol.XLVI, n°141, 73-86, 2008.

Révolution numérique et évaluation bibliométrique dans les sciences humaines et sociales

Christine Kosmopoulos

UMR Géographie-cités, Cybergeog

Denise Pumain

Université Paris I, Institut Universitaire de France.

Dans tous les domaines de l'activité sociale, les pratiques de l'évaluation tendent vers une meilleure explicitation, voire une plus grande transparence, de leurs objectifs et de leurs méthodes. Cette tendance n'épargne pas la production scientifique, dont le fonctionnement repose depuis très longtemps sur des méthodes d'évaluation par les pairs. Certaines disciplines ont commencé depuis plusieurs décennies à mettre au point des outils quantitatifs destinés à mesurer la qualité du travail des chercheurs et des revues en permettant la comparaison au moyen d'indicateurs chiffrés. Depuis une dizaine d'années, la « révolution numérique », qui se traduit non seulement par l'explosion du nombre des publications accessibles sous forme numérique, mais aussi par une profonde transformation des méthodes d'accès à l'information scientifique, semble devoir accélérer le recours à des méthodes bibliométriques pour l'évaluation, même dans le domaine des sciences humaines et sociales demeuré jusqu'ici très réticent à ce type d'approche.

La pratique de la bibliométrie trouve sa source dans le *Science Citation Index* qui répondait à l'origine à une demande documentaire hautement spécialisée, dans le domaine des sciences « dures » surtout, et à laquelle les professionnels n'étaient pas en mesure de répondre avec leurs instruments classiques. Cet outil appliqué strictement à l'analyse des articles scientifiques a progressivement évolué vers la bibliométrie. Or, en sciences sociales, non seulement les objets d'étude, les méthodes de recherche, le rapport au contexte social diffèrent de ceux des sciences dures, mais aussi les modes de communication scientifiques. Un rapport de Science-Metrix¹ sur "L'utilisation de la bibliométrie dans les sciences sociales et les humanités" montre à quel point dans ces disciplines, les modes de communication sont bien plus hétérogènes qu'en sciences dures (Archambault, Gagne, 2004). Alors que l'évaluation dans les sciences « dures » se fonde essentiellement sur les revues, les articles publiés dans des revues occupent moins de 50 % de la communication scientifique, les ouvrages, chapitres d'ouvrages, monographies, communications aux colloques, littérature grise, rapports occupant quant à eux plus de la moitié. Pour Hicks (1999), les ouvrages auraient un impact très important en sciences sociales, puisqu'ils comptabiliseraient 40 % des citations. Ces proportions, tout comme les pratiques de recherche et d'évaluation, varient d'ailleurs parmi les sciences sociales selon un spectre très large, des sciences économiques aux sciences juridiques en passant par la géographie par exemple.

¹ Structure canadienne spécialisée dans l'évaluation de la science auprès des pays anglo-saxons et scandinaves.

Dans le choix d'une base de citation de référence, du type de mesure bibliométrique, et de l'importance accordée à ces outils dans l'évaluation des recherches, les sciences sociales sont confrontées à des intérêts divergents. Les éditeurs privés tentent d'imposer leurs sources et les scientifiques déjà bien installés dans ces pratiques souhaitent diffuser leurs méthodes. Des stratégies adaptées aux outils de la révolution numérique sont déjà très visiblement en place, tandis qu'une alternative de publication en libre accès se profile. Les sciences sociales sauront-elles tirer parti de ce contexte encore incertain pour évoluer vers des pratiques mieux adaptées à leur statut ?

1 La bibliométrie : bases et indices

Les sciences sociales prises dans leur ensemble sont par profession méfiantes à l'encontre de toute forme de bibliométrie. Traitant par définition de sujets complexes, accoutumés aux pièges du comparatisme dans l'histoire ou entre les sociétés, très sensibles aux nombreux effets de contexte dont ils ont à rendre compte, les chercheurs en sciences sociales sont très sensibles aux effets de réduction qu'opère la quantification, notamment lorsqu'il s'agit d'observer par ce moyen une qualité scientifique. La mise en ordre institutionnelle de leurs savoirs et de leurs compétences par l'internationalisation des publications professionnelles est beaucoup moins avancée que dans le cas des sciences de la matière et du vivant. La diversité des cultures locales, la perte de sens associée aux traductions d'une langue à l'autre sont souvent mises en avant pour récuser toute tentative de dénombrement bibliométrique. Cependant, par facilité souvent, par calcul parfois, certaines institutions académiques ont commencé, dans certains pays même non anglophones, à utiliser la bibliométrie pour le classement ou le recrutement des chercheurs. Pourtant, les outils en vigueur sont encore bien imparfaits et souvent injustes dans le traitement qu'ils réservent aux recherches non anglophones et aux sciences sociales en général. Il importe donc de dresser un inventaire de ces outils pour permettre une analyse critique de leur utilisation. A plusieurs niveaux, celui des bases de données, celui des indices, et celui des moteurs de recherche, il est en effet possible d'intervenir pour que les outils bibliométriques rendent mieux justice à la production des chercheurs en sciences sociales.

1.1 Des bases de données bibliographiques et leurs enjeux

Trois grandes bases de données bibliographiques sont actuellement disponibles sur Internet. Etablies par des acteurs distincts, elles diffèrent par le nombre et le type de publications qu'elles recensent, qui ne se recouvrent que très partiellement et qui ne donnent chacune qu'une image très imparfaite de la production scientifique dans le champ des sciences sociales.

La première, la plus ancienne, connue aujourd'hui sous le nom de *Web of Science (WOS)*, a été constituée à l'origine par Eugène Garfield. Dès 1955, E. Garfield propose dans *Science* une idée nouvelle pour la communication scientifique : créer un index de citations des publications scientifiques. Pour l'auteur, un tel index répondait aux demandes d'information spécialisée des scientifiques, alors que le système classique d'indexation par les bibliothécaires offrait des entrées beaucoup trop généralistes et par conséquent inadéquates. L'objectif déclaré était hautement pertinent : recenser les sources citées dans un article, mais également les références corrélées, postérieures à sa publication. Il s'agissait là de mettre en évidence les réseaux des références scientifiques et de faciliter l'accès à l'information sur un thème donné, avec l'avantage pour les sciences dures d'accéder simultanément aux travaux de

la recherche appliquée et de la recherche fondamentale. Cette base, à l'origine intitulée « *Current Contents* » avait également pour mission déclarée d'aider les évaluateurs à vérifier la pertinence des sources bibliographiques des articles soumis aux revues (Garfield, 1964).

Mais l'idée répondait également à deux autres objectifs. Le premier consistait, par l'intermédiaire d'un recensement et d'une évaluation des contenus, à appliquer un contrôle sur la littérature scientifique mondiale. C'est ainsi que depuis sa création en 1964, le *Science Citation Index*, extrait des revues indexées dans la base des *Current Contents*, s'est imposé comme une source unique pour la citation des articles. Le second, d'ordre commercial, proposait un service de fournitures de tirés à part de tous les articles cités recensés dans les *Current Contents*. Cette entreprise privée, basée à Philadelphie, connue sous le nom de *International Scientific Institute* (ISI), est désormais la propriété du groupe Thomson. A ce jour, elle couvre pour l'ensemble des titres indexés une période de 60 ans, alors qu'un projet de numérisation de l'ensemble des documents existant depuis le début du 20^{ème} siècle est en cours.

Très récemment, en 2004, une seconde base, *Scopus*, a été instituée par le groupe d'édition Elsevier pour concurrencer Thomson Scientific. L'ambition de l'éditeur commercial est de permettre l'accès à l'ensemble de la littérature scientifique institutionnelle mondiale. Contrairement au *WOS*, il n'y a pas de commission d'évaluation ; mais *Scopus* indexe les articles des revues depuis 1996 selon une procédure quasi-similaire à celle du *WOS*. *Scopus* indexe les titres de revues qui lui sont soumis dès lors qu'ils sont labellisés par une institution. A ce jour, dans les deux cas, l'accès à ces bases de données privées est payant et se négocie (au prix fort) au niveau de la direction des organismes de recherche et des bibliothèques universitaires.

La troisième source, *Google Scholar*, lancée fin 2004, n'est pas à proprement parler une base de données bibliographique dans sa conception initiale ; il s'agit d'un moteur de recherche spécialisé dans la littérature scientifique. La base qui en est issue est la propriété du groupe qui a lancé le moteur de recherches *Google* en 1998. Les documents indexés proviennent des éditeurs scientifiques, des sociétés savantes, des dépôts d'archives ouvertes, des universités et d'autres organismes de recherche. *Google Scholar* utilise comme *Google* un outil robot qui vient explorer, classer et indexer le contenu d'un site qui lui a été signalé. Pour la France, *Google Scholar* prend en compte les archives déposées sur HAL (CCSD-CNRS)², une grande partie des fonds documentaires de l'INIST (13 millions de notices sur 20), les publications du portail Revues.org selon le protocole OAI.³

Ann Will Harzing, une universitaire australienne spécialiste d'informatique, a conçu de façon indépendante à partir de *Google Scholar* un outil de traitement et d'analyse des citations joyeusement intitulé « *publish or perish* » (2007)⁴. Contrairement aux deux bases précédemment citées, *Google Scholar* et *Publish or perish* sont en accès libre.

1.2 Les indices : outils d'évaluation

Les logiciels d'extraction d'information, qui permettent tout à la fois d'indexer ces publications en langage naturel et de calculer très rapidement le nombre de fois où elles sont citées, sont en effet la clé du succès grandissant de ces bases de données pour l'établissement

² <http://ccsd.cnrs.fr>

³ <http://revues.org>

⁴ Nous remercions Alain Peyraube de nous avoir signalé cette référence : <http://www.harzing.com/>

d'outils bibliométriques. Leurs performances vont bien au-delà de ce que permettaient les anciens systèmes de classification des bibliothécaires, le langage naturel se substituant de plus en plus aux classiques thesaurus ou aux listes de mots-clés, et l'indexation s'effectuant désormais non seulement par nom d'auteur, titre de la revue et de l'ouvrage, date de publication et références éditoriales, ainsi que le titre de l'article ou de l'ouvrage, mais aussi sur l'ensemble du texte publié (*full text indexation*).

Les logiciels de *fouilles de données* (*data mining*), qui comme pour *Google Scholar* explorent ce gigantesque corpus se limitent parfois à comptabiliser le nombre des citations d'un article, trouvées dans les bibliographies d'autres articles ou ouvrages. C'est ce que proposait à l'origine le « citation index » de l'ISI. Mais le plus souvent, des outils de calcul plus perfectionnés ont été intégrés à cette recherche et en permettent, jusqu'à un certain point, l'analyse. Les outils proposés s'appuient sur des études récentes analysant l'explosion des observations scientifiques en bibliométrie⁵, les outils d'exploration et la mise au point d'indices significatifs ayant beaucoup progressé de conserve au cours des dix dernières années.

Les indices construisent de la valeur pour les revues et les auteurs et chercheurs, en se fondant sur un *a priori* simple : le nombre de fois où un article ou un ouvrage est cité témoigne de son utilité et de sa valeur scientifique. Ce point peut évidemment être discuté, un nombre de citations important pouvant aussi bien refléter des controverses, des réfutations, ou simplement un effet de mode, mais en moyenne il peut être considéré comme significatif d'une valeur de réputation. Le nombre de citations par auteur permet d'établir le *coefficient d'impact* d'une revue, qui témoigne de son audience et de sa capacité à sélectionner les meilleurs auteurs. Plusieurs revues spécialisées, comme *Scientometrics*, *Infometrics*, le *Journal of Information Science*, ont été créées pour traiter des questions relatives à la mesure de la valeur des publications scientifiques.

Dès 1935, une « loi » établie par Bradford montrait que cette valeur de réputation est très inégalement distribuée, 150 revues scientifiques sur plusieurs milliers comptabilisant la moitié du nombre total des citations. Lorsque les articles et ouvrages d'un auteur sont rangés par ordre décroissant de leur nombre de citations, cette série commence en général par un nombre relativement élevé, de quelques dizaines, voire centaines ou milliers de citations pour une référence très souvent citée, mais ce nombre diminue ensuite assez rapidement pour les écrits suivants d'un même auteur. La série des nombres de citations se rapproche davantage d'une progression géométrique (ou d'une exponentielle) décroissante que d'une diminution linéaire (arithmétique) du nombre de citations. Ainsi, puisque la forme générale de la distribution statistique du nombre de citations par article ou ouvrage, pour un même auteur, n'est pas symétrique, mais au contraire très dissymétrique, de nombreux articles ne recevant aucune citation ou très peu, pour quelques-uns qui en reçoivent beaucoup, le nombre qui peut représenter le mieux cette distribution n'est pas une moyenne de citations par article, qui caractériserait une distribution normale du nombre de citations, mais un nombre qui tienne compte de la forme dissymétrique de cette distribution. Le physicien Hirsch (2005) a proposé de retenir le nombre *h*, égal au rang de l'article qui a reçu au moins autant de citations que la valeur numérique de son rang. Ce nombre matérialise en quelque sorte l'intersection de la courbe représentant la progression géométrique inverse du nombre des citations avec la

⁵ Manuel Durand-Barthez, *Citations et Facteurs d'impact : quel avenir pour l'évaluation ?* en ligne sur le site de l'URFIST : http://urfistinfo.blogs.com/urfist_info/2006/04/citations_et_fa.html

progression arithmétique (linéaire) croissante du rang des articles. La position de ce nombre est représentative de l'ordre de grandeur de la série du nombre des citations. Mathématiquement, ce nombre est proche de celui du logarithme du nombre des citations – mais cette dernière mesure serait moins directement « parlante » pour les utilisateurs de l'indice.

Bien que de création relativement récente -2005- cet indice a été considéré comme suffisamment robuste et explicite pour être intégré dès l'année suivante dans les calculs bibliométriques diffusés par Thomson/ISI sur le *Web of Science*.

1.3 Une couverture insuffisante pour les SHS

Les trois grandes bases actuelles de citations de la littérature scientifique présentent chacune des lacunes importantes, qui ne permettent pas une évaluation pertinente des chercheurs ou des revues (*impact factor*) au moyen des outils bibliométriques

Bien que considérée par certains comme la base de citations de référence, le *Web of Science* est critiqué depuis de nombreuses années pour sa partialité à l'encontre des SHS. Il recense quelque 8700 revues internationales (on estime à environ 20000 le nombre de revues scientifiques dans le monde), mais seulement 1000 pour les SHS, dont la quasi-totalité est d'origine anglo-saxonne. A titre d'exemple, les revues SHS soutenues par le CNRS ne sont pas présentes dans le WOS. Le CNRS a tenté de négocier en 2004 des améliorations auprès de Thomson/ISI qui prendraient en compte les besoins et les spécificités européennes, mais il n'y a pas eu d'avancée. Un rapport remis en 2003 par Philippe Jeannin au ministère de la recherche et des nouvelles technologies, sur l'évaluation de la recherche en SHS, confirme l'absence de couverture des revues françaises dans ces domaines.

En vue de mieux analyser la production scientifique européenne dans le WOS, la *European Scientific Foundation (ESF)* s'est lancée dans l'évaluation des revues en SHS et a publié en juin 2007 une liste initiale sous le nom de *European Reference Index for the Humanities (ERIH)*⁶, qui les classe par rang (A,B,C). Cette liste ne comprend pas certains domaines des sciences sociales, comme la géographie, qui feront l'objet d'un autre classement. L'ESF statue sur l'intégration des revues dans cette liste de référence, ce qui explique pourquoi toutes les revues soutenues par le CNRS n'y figurent pas. Mais si l'*ERIH* n'a pas vocation à être un outil bibliométrique et ne se pose pas comme une alternative au WOS, l'*Observatoire des Sciences et des Techniques (OST)* qui a précisément pour mission de concevoir et de produire des indicateurs sur la recherche et le développement, annonce qu'il va travailler sur les indices d'impact à partir des revues que recense l'*ESF*⁷.

La base *Scopus* (Elsevier), pourrait être une alternative au monopole de Thomson. En effet, *Scopus* indexe environ 17000 titres dont 2850 en SHS, soit le double du WOS, et ne se limite pas aux revues anglo-saxonnes. La répartition géographique des titres est de 25% pour le Royaume-Uni (4157 revues), 25% seulement pour le reste de l'Europe/Moyen-Orient/Afrique, 37% pour l'Amérique du Nord, 12% pour Asie/Pacifique, et 1% pour l'Amérique du Sud. La couverture est donc beaucoup plus large que celle du WOS. De plus, les requêtes dans cette base *Scopus*, associée au moteur de recherche Scirus, apportent un

⁶ <http://www.esf.org/research-areas/humanities/activities/research-infrastructures/faq-sheet/scope-initial-lists.html#c13190>

⁷ <http://www.obs-ost.fr/>

éventail de résultats beaucoup plus large car il comprend aussi d'autres types de documents que les articles publiés dans les revues. La liste des sources de Scirus est indiquée sur le site. Malheureusement, la période couverte (11 ans) par Scopus reste encore très limitée. Une étude comparative du WOS et de *Scopus* pour une dizaine de disciplines en SHS est actuellement en cours sous la direction de Christine Kosmopoulos.

En fin de compte, dans les deux cas, la couverture des publications dans les revues, tant disciplinaire que temporelle, est tellement incomplète qu'elle donne une représentation biaisée de la production scientifique en SHS et ne peut par conséquent pas fournir des résultats solides pour une évaluation bibliométrique.

Hormis sa gratuité, *Google Scholar* a l'avantage de recenser toute la littérature scientifique sans distinction : les articles dans des revues, qu'elles soient ou non à comité de lecture, mais aussi des thèses, livres, extraits d'ouvrages, rapports, pré-prints etc. Toutefois, ce moteur de recherche spécialisé pose d'autres problèmes. Contrairement au WOS ou à *Scopus*, il ne donne aucune information sur les ressources qu'il utilise. Comme l'indique Jean-Pierre Lardy dans sa publication en 2007 sur le site de l'URFIST⁸, on ne trouve aucune liste des éditeurs commerciaux et des serveurs d'archives indexés, aucun renseignement sur la période couverte, le volume ou encore les pays concernés. Certaines sources connues de *Google Scholar*, comme par exemple la base bibliographique *Francis* de l'INIST, ne répondent pas aux normes bibliométriques. La couverture des archives des éditeurs comporte également d'importantes lacunes. Un test sur des personnalités scientifiques des sciences vient confirmer cette observation comme le montre Peter Jacso⁹. Une révision importante et systématique serait nécessaire pour que cette base devienne une référence incontestable pour les citations.

L'opacité des sources et l'absence d'exhaustivité (dans une même revue, toutes les publications d'un même auteur n'apparaissent pas nécessairement) des informations extraites par l'outil *Google Scholar* et analysées par Anne-Wil Harzing sur son site¹⁰, n'apportent pas à ce jour la garantie d'une évaluation fiable. Cependant les travaux de A.-W. Harzing permettent de faire de cette base un usage critique¹¹. Le fait est qu'à ce stade, l'outil "*Google Scholar-Harzing*" assure une bien meilleure prise en compte des diverses formes de communication en SHS que le WOS ou *Scopus*. En outre, son principal avantage est qu'il permet d'intervenir dans les calculs d'indice, en sélectionnant les publications jugées réellement significatives de l'activité scientifique d'un auteur, en éliminant les publications d'auteurs homonymes, les doublons ou les références peu pertinentes, et en proposant toute une variété d'indices qui corrigent l'indice h. Ainsi, l'indice g (*Egghe's g-index*) est calculé sur le même principe que l'indice h mais donne plus de poids aux articles très cités. Un autre indice (*Individual h-index*) corrige l'indice h par le nombre moyen d'auteurs par article, afin de permettre la comparaison des productions entre des disciplines où les pratiques d'identification des auteurs d'un article sont très différentes. Un autre complément de l'indice h est sa pondération par le nombre d'années d'activité de publication de l'auteur (*Age-weighted citation rate*), ce qui autorise à comparer des personnes qui sont à différents stades dans la carrière scientifique. L'indice h peut également être calculé en donnant plus de poids aux articles récents, en attribuant un score plus élevé aux personnes qui continuent à être productives (*Contemporary h-index*).

⁸ <http://urfist.univ-lyon1.fr/GoogleScholar.pdf>

⁹ <http://www2.hawaii.edu/~jacso/extra/gs/>

¹⁰ http://www.harzing.com/resources.htm#pop_gs.htm

¹¹ <http://www.harzing.com/>

Dans la mesure où la base de données *Google Scholar* couvre mieux les publications en sciences humaines et sociales, et notamment les ouvrages, il serait sans doute utile d'entreprendre, pour chaque discipline et sous-discipline, un étalonnage systématique de ces différentes mesures. A.W. Harzing y invite d'ailleurs sur son site. La valeur des indices peut être considérablement enrichie par une meilleure connaissance de leur qualité de représentation des valeurs scientifiques habituellement reconnues par les pairs. En effet, en dépit des réticences quant à leur usage et de l'imperfection des bases de données existantes, il est très probable que l'expansion des publications en ligne et la facilité de leur consultation conduise, à brève échéance, à l'adoption des indices de citations dans les procédures institutionnelles de recrutement et d'évaluation des chercheurs. Plutôt que de se laisser imposer les instruments considérablement biaisés du *Web of Science* qui sont actuellement utilisés pour nombre de disciplines des sciences de la matière et du vivant, les chercheurs des sciences humaines et sociales auraient sans doute intérêt à s'approprier des outils fondés sur une base de données plus ouverte à leurs pratiques de publication. Certes, toutes ces mesures quantifiées ne sauraient se substituer complètement à l'évaluation plus qualitative, en partie subjective, mais bien plus subtile et assurée, que représente la classique évaluation par les pairs. Même dans les disciplines expérimentales, où il est généralement admis que le nombre des citations reflète assez bien la qualité de la production d'un scientifique, et où le classement des revues est systématiquement pris en compte dans les stratégies de publication, les outils de bibliométrie font l'objet de critiques récurrentes. En fait, c'est la pratique sociale de la citation scientifique qui mérite un examen critique.

2 La citation est un sport de combat

Tout auteur scientifique sait bien que la liste des références qu'il cite dans un article à l'appui d'un travail donné est en partie arbitraire et souvent lacunaire. Outre l'impossibilité objective de connaître et citer l'ensemble des références utiles pour un sujet donné, le biais principal est sans doute disciplinaire, chaque science ignorant délibérément les avancées des autres, et cela d'autant plus qu'elle se situe à un rang élevé de la hiérarchie implicite des valeurs d'usage, (ou plutôt de la valeur d'intérêt) qui mettrait par exemple l'économie tout en haut du panier et les humanités tout en bas. Mais les orientations disciplinaires ne sont pas seules en cause, et, dans des métiers où la principale rémunération est un capital de réputation, la citation est aussi une arme stratégique. Pour quelques références « incontournables », combien de citations de complaisance, d'oublis délibérés, de renvois d'ascenseurs internes dans des « chapelles » qui s'ignorent réciproquement, de coquetteries de la citation « exotique » (c'est l'une des explications faciles du succès de la « *French Theory* » dans les études culturelles d'outre Atlantique ou d'outre Manche) ou encore d'accrochage indu à un auteur de réputation majeure ?

Ce n'est pas ici le lieu de même seulement esquisser une sociologie ou une psychologie sociale de l'activité de citation scientifique. On voudrait simplement souligner à quel point les distorsions habituelles de ces pratiques, qui se trouvent entérinées par une mesure « objective » de réputation via les citations, sont encore amplifiées par la mise en ligne et la mondialisation de l'accès aux références. De nouvelles stratégies adaptées à ces media sont apparues. On connaît les manipulations d'affichage que permettait le procédé initial de recherche de Google, fondé sur des mots-clés, qu'il suffisait d'introduire en grand nombre dans des pages cachées pour faire émerger un site dans les premières places. Depuis, des algorithmes procédant à partir des réseaux sémantiques ont été mis en œuvre, qui rendent plus complexes les moyens d'acquérir la visibilité et se faire citer, sans toutefois les empêcher. En effet, la proximité des ressources d'information disponibles sur le web, que les chercheurs

emploient aussi pour réaliser leurs propres recherches bibliographiques, sans nécessairement se limiter à leurs revues scientifiques habituelles, renforce la visibilité de « produits » scientifiques lorsqu'ils sont aussi très souvent cités par d'autres sources.

Parmi les possibilités de manipulation du système bibliométrique, citons celle qui consiste à faire savoir, en tant qu'évaluateur d'articles, que l'on acceptera tout article citant l'une de ses propres publications (mais cela n'est peut-être qu'une rumeur...). Une autre stratégie consiste à multiplier les domaines de citation possibles, en jouant sur des usages du vocabulaire aux frontières du sens commun ou des attentes du grand public, ce qui est fréquent dans les sciences humaines et sociales. Une partie du succès foudroyant de l'auteur de l'expression « *creative class* » (catégorie sociologique « nouvelle » qui englobe les créateurs de tout bords, artistes comme entrepreneurs), pour leur attribuer la responsabilité du dynamisme des villes par l'innovation, s'explique sans doute par une exploitation habile de ces amalgames. L'effet de reprise par d'autres médias amplifie aussi la visibilité des articles scientifiques qui traitent de « sujets de société », comme les homosexualités ou encore les *gender studies*. La collusion médiatique, volontaire ou non, est fréquente lorsque les chercheurs travaillent sur des sujets qui rencontrent parfois l'actualité, qu'il s'agisse de risques, de crises, de criminalité, de terrorisme ou même de simple politique, notamment en période électorale. Le courant post-moderne, qui a milité pour l'introduction du savoir profane dans les sciences humaines et sociales, en reproduisant des « paroles d'habitants » ou en légitimant le savoir contenu dans les discours tenus par les personnes observées, contribue aussi à brouiller les frontières, lorsque les liens et mots-clés sur le web accrochent aussi bien les divers blogs que les articles en ligne.

Les bulles spéculatives vont-elles éclater ou bien laisser des traces ? Observera-t-on sur le web une évolution de type biologique, par « sélection naturelle », avec un temps de décanation qui laisserait surnager les écrits les plus aptes par leur qualité à nourrir la soif de savoir des internautes, ou bien ne voit-on pas déjà plutôt s'installer un processus évolutif sociétal, selon lequel « la mauvaise monnaie chasse la bonne » ? Il semble que, y compris sur la toile, les « factoides » aient la vie dure et que la corrélation entre visibilité médiatique et qualité scientifique ne se renforce guère, justifiant le pessimisme qu'exprime par exemple Andrew Keen (2007). Aura-t-on un jour les moyens techniques de détecter les auteurs de citation en rond, d'apporter des correctifs aux réseaux « commerciaux » sémantiques ou paradigmatiques, au moyen des mêmes outils performants de l'analyse sémantique sur le web ? En attendant, une intéressante garantie de qualité est offerte par l'intervention de nouvelles pratiques dans la production en ligne, sous le contrôle d'une évaluation par les pairs. Elle implique que l'édition scientifique puisse se dégager des intérêts commerciaux, grâce à une valorisation « en interne » de l'expertise des chercheurs.

3 Pour un changement de régime de valeurs

La communication scientifique utilise désormais les nouvelles technologies de la communication et de l'information (NTCI), ce qui a pour effet constaté une réforme profonde des modes de diffusion de l'information scientifique. La plupart des travaux scientifiques sont numérisés et mis en ligne, soit par les chercheurs¹², soit par les éditeurs commerciaux¹³, sur des sites d'archives institutionnelles ouvertes¹⁴, sur des sites publics¹⁵ ou encore sur des pages

¹² On prend pour exemple le portail de revues.org du CNRS

¹³ Cairn (consortium d'éditeurs français), Elsevier, Springer Verlag etc.

¹⁴ Pour la France, le CCSD du CNRS

personnelles. Les NTCI modifient donc les comportements des chercheurs qui ont désormais les moyens d'augmenter la visibilité de leurs travaux à l'échelle mondiale et d'accroître leur potentiel de citabilité par la mise en ligne, mais elles marquent aussi un tournant important à la fois dans les relations avec les éditeurs scientifiques commerciaux et dans les pratiques de travail. En effet, le développement des NTCI peut nous faire entrer dans une ère nouvelle, une ère du partage c'est-à-dire du *Libre Accès*. Mais cette évolution est contrée par la résistance des acteurs financiers, comme en témoigne la création toute récente par l'*Association des éditeurs américains* de l'organisation anti-Open Access *PRISM (Partnership for Research Integrity in Science & Medicine)*.

3.1 Le partage des ressources

Depuis quelques années un bras de fer s'est enclenché entre les éditeurs commerciaux, qui revendent aux chercheurs/auteurs leurs publications dont ils ont acquis les droits, et le mouvement international de l'*Open Access* (OA), officialisé par l'appel de Budapest en décembre 2001, qui revendique le libre accès à tous les documents scientifiques (écrits, thèses, articles scientifiques etc.)¹⁶. Près de 400 universités et instituts scientifiques de par le monde y adhèrent.¹⁷ De nombreuses initiatives vont dans ce sens, comme par exemple le projet *Sherpa*¹⁸ en Angleterre, dont la mission est de développer des réservoirs de documents numériques universitaires en accès libre ou encore *Driver*, une infrastructure européenne de soutien aux réservoirs numériques en libre accès¹⁹. Pour avoir une idée des ressources disponibles en OA, on peut se référer à *OpenDoar*, répertoire mondial des dépôts en archives ouvertes²⁰ ou le *Registry of Open Access Repositories (ROAR)*²¹.

L'*Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH)*²² s'inscrit dans le mouvement des *Archives Ouvertes* ou *Libre Accès*²³. Son objectif est de rendre inter-opérables les métadonnées sur les documents en ligne, afin de permettre le transfert des données entre les différents serveurs qui respectent ce protocole. Le principe est simple, les données sont déposées sur le serveur dans un entrepôt et sont rendues accessibles à tous les moissonneurs OAI qui naviguent (et glanent !) sur la toile. Des centaines de sites respectent désormais les normes OAI-PMH comme par exemple *Gallica* de la Bibliothèque Nationale de France, *Sudoc* (France), *Eprints* (Angleterre), la *bibliothèque du Congrès* (Etats-Unis) etc. Le moteur de recherche scientifique *OAIster* de l'Université du Michigan moissonne quant à lui dans 714 bases de données mondiales et recense plus de 13 millions de références.

L'idée du libre accès aux publications scientifiques, qu'incarne l'appel de Budapest, est défendue depuis de nombreuses années par les chercheurs ; elle s'est concrétisée lors de la mise en ligne du premier serveur de dépôt des archives ouvertes à Los Alamos, *ArXiv*, par Paul Ginsparg en 1991. En 2001, le *Centre de la Communication Scientifique Directe (CCSD)* du CNRS lance à son tour une archive multidisciplinaire internationale ouverte, fondée sur

¹⁵ *Persée* (ministère de l'éducation nationale, de l'enseignement supérieur et de la recherche) procède à une numérisation rétrospective des revues en SHS.

¹⁶ <http://www.soros.org/openaccess>

¹⁷ <http://www.soros.org/openaccess/view.cfm>

¹⁸ <http://www.sherpa.ac.uk/index.html>

¹⁹ <http://www.driver-support.eu/en/index.html>

²⁰ <http://www.opendoar.org/index.html>

²¹ <http://roar.eprints.org/>

²² <http://www.openarchives.org/>

²³ *Open Access* étant souvent traduit de l'une ou l'autre façon. Il comprend à la fois les publications en libre accès, mais aussi les serveurs de dépôt d'archives ouvertes.

l'auto-archivage libre. Respectueuse du protocole OAI-PMH, elle permet des échanges entre les grandes bases de données scientifiques internationales et les moteurs de recherche²⁴.

Hormis les archives ouvertes qui comprennent aussi les sites de dépôt de thèses, le nombre de revues scientifiques électroniques en accès libre produites par les chercheurs va croissant. 2811 sont recensées dans le DOAJ²⁵. Dans le domaine des SHS, la plus ancienne revue électronique, *Cybergeo : European Journal of Geography*²⁶, existe depuis 1996. Reconnue par le CNRS, elle comporte un comité de lecture international et publie dans toutes les langues européennes. Produite par des universitaires et des chercheurs, indépendante de tout éditeur commercial et libre d'accès, *Cybergeo* est recensée dans *Scopus*, dans *Google Scholar* via le portail de Revues.org auquel elle adhère depuis 2007 et qui applique le protocole OAI-PMH.²⁷ D'autres initiatives sont à noter comme *Hypergeo, encyclopédie plurilingue de géographie*²⁸ produite par des universitaires à l'intention des enseignants et des étudiants, dont l'objectif est de présenter les principaux concepts et théories de la géographie.

3.2 Les plates-formes collaboratives en Libre Accès

Sous l'impulsion de l'*Open Access Initiative*, le partage prend donc de multiples facettes : archives ouvertes, publications en accès libre, protocole OAI-PMH, mais aussi des outils collaboratifs, logiciels libres (*Open Source*), droits d'auteurs de type *Creative commons*. La plupart des plates-formes de gestion de ressources numériques proposées en "open source" telle que *DSpace*²⁹, permettant d'archiver, d'indexer et de diffuser des contenus numériques, sont compatibles avec le protocole OAI-PMH.

Pour le travail collectif et l'auto-publication des documents, de nombreuses applications sont disponibles en Libre Accès, on notera *Google Document et Tableur* qui permet de créer ou d'importer des documents et de travailler à plusieurs auteurs. Des sites de partage de photos comme *Flickr* sont déjà utilisés par certains laboratoires pour gérer et publier leurs fonds. Dans la même idée, les Wikis qui permettent de créer, de modifier et de publier du contenu de manière collaborative, le plus connu étant évidemment Wikipédia où chacun peut interagir et enrichir les pages. L'accès à la visualisation ou l'édition des pages peut toutefois être limité par un mot de passe, c'est le cas de wikis scientifiques organisés par des groupes de travail sur un sujet donné.

D'autres types de plates-formes de partage ont fait leur apparition qui marquent bien le tournant amorcé et la révolution en cours dans les modes de travail. Les agrégateurs en ligne³⁰, par exemple, offrent la possibilité de partager les flux de syndication³¹ si l'abonné le souhaite. Concrètement, en s'abonnant au flux RSS de *Cybergeo*, on voit s'afficher le nombre

²⁴ OAISter, Google Scholar etc.

²⁵ <http://www.doaj.org/>

²⁶ <http://www.cybergeo.eu>

²⁷ *Cybergeo* n'a jamais été acceptée par le comité d'évaluation du WOS bien que la revue réponde à tous les critères requis. L'argument avancé est que l'impact factor de la revue dans le recensement de la base de données du WOS est nul. Or aussi longtemps qu'une revue n'est pas intégrée dans le WOS, elle ne peut y être citée. Le message est suffisamment clair. En réalité, une revue scientifique en SHS, de surcroît francophone et qui plus est en accès libre n'intéresse pas plus Thomson, qu'elle n'intéressait ISI auparavant.

²⁸ <http://www.hypergeo.eu/>

²⁹ <http://www.dspace.org/> : logiciel de répertoire numérique et d'archivages à l'intention des universités et organismes de recherche

³⁰ Exemples : bloglines, lamooche, RSS-Feed etc.

³¹ Un fil RSS ou flux de syndication permet de collecter les nouveautés mises en ligne via un agrégateur, en remplacement des lettres d'informations ou des systèmes d'alerte via le web mail.

de personnes abonnées à ce flux ; en cliquant sur la liste, les profils défilent, on peut alors accéder à l'interface de l'un des profils sélectionné et lire, utiliser ou rapatrier dans son propre agrégateur les flux d'information que la personne a recueillis.

Il en est de même pour les favoris (Internet Explorer) ou marque-pages (Firefox). Au lieu de conserver ses marque-pages dans l'interface locale de son navigateur, on peut décider de les collecter sur une plate-forme de partage comme par exemple *del-icio-us* ou *Google Reader*, avec pour avantage de disposer à partir de n'importe quel poste des connexions aux pages taggées et simultanément de pouvoir échanger les données. *Connotea* fonctionne également sur ce principe, mais apporte en plus un service très utile de gestion des références et des articles en ligne, conçu tout spécialement à l'intention des scientifiques. Comme pour les flux RSS, on peut décider de rendre visibles ses propres marque-pages, classifications, références avec accès au texte intégral, et accéder à celles des autres.

4 Conclusion : penser la bibliométrie dans le contexte des NTCI

La position des sciences sociales n'est pas bien assurée dans les systèmes d'analyse bibliométrique, pour des raisons liées à des différences de pratiques scientifiques, qu'on a parfois tendance à imputer à un présumé « retard » de ces disciplines par rapport à ce qui s'est mis en place dans les sciences dites « dures », mais qui tiennent aussi probablement d'autres facteurs, notamment la diversité culturelle et linguistique et l'évolution historique plus rapide des sujets de recherche. Ceux-ci sont aussi des sujets « de société » impliquant de fortes interactions entre l'objet de science et le contexte social dans lequel il est analysé (Latour, 1996). D'aucuns appuient d'ailleurs leur rejet de toute bibliométrie en sciences humaines sur cette différence supposée irréductible des « sciences historiques » (l'expression est de Jean-Claude Passeron, 1991).

Le développement des outils collaboratifs, grâce à l'Internet de 2^e génération (web 2.0) et le mouvement du *Libre Accès*, l'avancée dans le domaine de l'internationalisation des normes d'échange, offrent des bases nouvelles à la réflexion sur la bibliométrie en général et ouvrent des perspectives pour améliorer ses usages en sciences humaines et sociales. Les moyens techniques permettent désormais une diffusion et un partage des travaux scientifiques pour un coût modique et un accès de plus en plus large aux ressources, notamment via le protocole OAI-PMH ou les flux RSS, avec pour conséquence l'augmentation du bassin de citations, indispensable pour parvenir à des résultats d'analyse solides. On pourrait imaginer la constitution de nouveaux grands répertoires ou de nouvelles bases de données qui combindraient les bases de données existantes (moteurs de recherche scientifiques, archives ouvertes, bibliothèques etc.), ainsi que les références bibliographiques fournies sur les pages web des chercheurs. On pourrait également ajouter à cette base de données statistiques un indicateur à partir de l'opinion des pairs (Raaijmakers, 2003).

L'expérience menée de 1999 à 2002 par les universités de Southampton, Cornell et arXiv.org, *Open Citation Project*³², aboutit à la conclusion que la recherche des rangs de citations et des liens entre les références est un des exemples des services OAI, elle encourage les chercheurs à déposer leurs travaux sur les sites institutionnels. La base de données HAL lancée par le CCSD du CNRS fait évidemment partie de ce projet et pourrait servir d'outil bibliométrique. Depuis début 2006, le CNRS a entamé une grande campagne d'information allant dans ce sens. L'ESFRI (*European Strategy Forum on Research Infrastructures*) réfléchit sur de

³² <http://opcit.eprints.org/>

nouvelles formes d'évaluation des SHS dans la perspective du 7ème programme-cadre (7e PC) de recherche de l'Union européenne³³. La Fondation Européenne pour la Science a entrepris le recensement et la validation de listes de revues pour chaque domaine des sciences humaines et sociales, avec un classement selon leur réputation.

Il reste que, dans le domaine de la bibliométrie comme dans celui de l'évaluation en général, se fier à un indicateur unique, si sophistiqué soit-il, reste une ambition totalement inadaptée à ce que l'on sait de la complexité des systèmes sociaux, et qu'il convient de mettre au point, non seulement des batteries d'indicateurs, mais aussi des méthodes multivariées pour leur analyse et pour la préparation de toute décision. La mise en réseau de la recherche à l'échelle internationale grâce au support et à la communication électroniques devrait faire émerger de nouvelles formes de l'évaluation scientifique, mieux harmonisées, dont les outils de bibliométrie ne sont qu'un aspect.

Bibliographie :

Archambault E., Gagne E., 2004 (dir), *L'utilisation de la bibliométrie dans les sciences sociales et les humanités*, Science -Metrix, Rapport final préparé pour le Conseil de recherches en sciences humaines du Canada (CRSH) : http://www.science-metrix.com/fr/projets_2004_Bibliometrie_Sciences_Sociales.htm

Aubry C., Janik J. (dir), 2005, *Les archives ouvertes. Enjeux et pratiques*, Paris, ADBS éditions.

Comité national d'évaluation de la recherche (CNER), 2006, Pour une meilleure évaluation de la recherche publique en sciences humaines et sociales, 72 p.: <http://www.cner.gouv.fr/cgi-bin/htsearch>

Garfield E., 1955, "Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas", *Science*, 122(3159), 108-111.

Garfield E., 1964, "Science Citation Index -- A New Dimension in Indexing", *Science*, 144(3619) 649-654.

Harzing 2007, Reflections on the h-index. http://www.harzing.com/pop_index.htm.

Hicks D., 1999, "The difficulty of achieving full coverage of international social science literature and the bibliometric consequences", *Scientometrics*, 44(2), 193-215.

Keen, Andrew 2007. *The Cult of the Amateur: How the Democratization of the Digital World is Assaulting Our Economy, Our Culture, and Our Values*. Doubleday, Currency.

Jeannin Ph, 2003, *Revue-métrie de la recherche en SHS* - Rapport synthétique et final de mission (1999-2003), Ministère délégué à la recherche et aux nouvelles technologies, France : http://www.revues.org/cost/images/1/12/JEANNIN_-Rapport_final2003a.pdf

Latour B., 1996, *Petites Leçons de sociologie des sciences*, Paris, Le Seuil.

Passeron J.C. 1991, *Le raisonnement sociologique. L'espace non-poppérien du raisonnement naturel*. Paris, Nathan.

Van Raan, A.F.J., 2003, The use of bibliometric analysis in research performance assessment and monitoring of interdisciplinary scientific developments. *Technikfolgenabschätzung, Theorie und Praxis* 12(1), 20-29.

³³ <http://cordis.europa.eu/esfri/> et http://cordis.europa.eu/fp7/home_fr.html